Preregistration

Ranking task with three types of event sets

Xiaotong Liu[1], Arndt Bröder[1], Henrik Singmann[2]

[1] University of Mannheim

[2] University College London

10. February 2022

Study Information

Title

Ranking task with three types of event sets

Description

Various sampling-based models have been proposed to account for biased probability judgments. The core assumptions underlying these models are that people take samples from coherent sets of probability distributions in mind and then work with samples to generate explicit judgements. To date, sampling-based models of probability judgements have only been investigated in the probability estimation task (e.g., Costello & Watts, 2014).

Our goal is to use an event ranking task to investigate sampling-based models of probability judgements. On each trial of the ranking task, participants will be asked to provide a ranking for an event set consisting of two pairs of complementary events, $\{A, \neg A, B, \neg B\}$ (where $\neg$ denotes "not"), regarding their perceived likelihoods of occurrence. There is a finite number of answers that a participant can give when asked to create a ranking for four events. When ties are prohibited in the rankings, there are 24 possible rankings over four events. 24 rankings can be divided into three categories: Type-1 ranking, Type-2 ranking and Logically possible ranking (see Table 1 for details). When ties are allowed in the rankings, there are 75 possible rankings over four events. 75 rankings can be divided into four categories: Type-1 ranking, Type-2 ranking, Type-3 ranking and Logically possible ranking (see Table 1 for details).

We formally model the process underlying how people generate a ranking from sampling, assuming direct sampling from the fixed underlying probability distributions that follows a binomial process. We also add read-out noise as suggested in one existing sampling-based model, the Probability Theory plus Noise (PT+N) model (Costello & Watts, 2014) in the sampling process. We assume that three aspects, each controlled by unique parameters, determine the probability of giving different rankings:

- An array of four fixed probabilities, $\{P(A), 1 - P(A), P(B), 1 - P(B)\}$, which quantifies the inherent beliefs regarding the probabilities of events in an event set,
- the error rate, d, which quantifies the probability of mistaking an instance as its complement when reading a sample, and
- the sample size, N, which quantifies the number of instances sampled from an internal distribution.

In a simulation, we derived qualitative predictions for probabilities of giving different types of rankings by manipulating inherent beliefs. The qualitative patterns are almost irrespective of the sample size N and error rate d. The current study aims to test the qualitative predictions we derived via simulation.

	Category	Ranking:			
		1st place	2nd place	3rd place	4th place
1		A	B	$\neg B$	$\neg A$
2		A	$\neg B$	B	$\neg A$
3		$\neg A$	B	$\neg B$	A
4		$\neg A$	$\neg B$	B	A
5		B	A	$\neg A$	$\neg B$
6		B	$\neg A$	A	$\neg B$
7		$\neg B$	A	$\neg A$	B
8	Logically possible	$\neg B$	$\neg A$	A	B
9	ranking	A	B, $\neg B$	$\neg A$	
10		$\neg A$	B, $\neg B$	A	
11		B	A, $\neg A$	$\neg B$	
12		$\neg B$	A, $\neg A$	B	
13		A, B	$\neg A$, $\neg B$		
14		A, $\neg B$	$\neg A$, B		
15		$\neg A$, B	A, $\neg B$		
16		$\neg A$, $\neg B$	A, B		
17		A, $\neg A$, B, $\neg B$			
18		A	$\neg A$	B	$\neg B$
19		A	$\neg A$	$\neg B$	B
20		$\neg A$	A	B	$\neg B$
21		$\neg A$	A	$\neg B$	B
22		B	$\neg B$	A	$\neg A$
23		B	$\neg B$	$\neg A$	A
24		$\neg B$	B	A	$\neg A$
25		$\neg B$	B	$\neg A$	A
26	Type-1	A, $\neg A$	B	$\neg B$	
27	ranking	A, $\neg A$	$\neg B$	B	
28		B, $\neg B$	A	$\neg A$	
29		B, $\neg B$	$\neg A$	A	
30		A	$\neg A$	B, $\neg B$	
31		$\neg A$	A	B, $\neg B$	
32		B	$\neg B$	A, $\neg A$	
33		$\neg B$	B	A, $\neg A$	
34		A, $\neg A$	B, $\neg B$		
35		B, $\neg B$	A, $\neg A$		
36		A	B	$\neg A$	$\neg B$
37		A	$\neg B$	$\neg A$	B
38		$\neg A$	B	A	$\neg B$
39		$\neg A$	$\neg B$	A	B
40		B	A	$\neg B$	$\neg A$
41		B	$\neg A$	$\neg B$	A
42		$\neg B$	A	B	$\neg A$
43		$\neg B$	$\neg A$	B	A
44		A, B	$\neg A$	$\neg B$	
45		A, B	$\neg B$	$\neg A$	
46		A, $\neg B$	$\neg A$	B	
47	Type-2	A, $\neg B$	B	$\neg A$	
48	ranking	$\neg A$, B	A	$\neg B$	
49		$\neg A$, B	$\neg B$	A	
50		$\neg A$, $\neg B$	A	B	
51		$\neg A$, $\neg B$	B	A	
52		A	B	$\neg A$, $\neg B$	
53		A	$\neg B$	$\neg A$, B	
54		$\neg A$	B	A, $\neg B$	
55		$\neg A$	$\neg B$	A, B	
56		B	A	$\neg A$, $\neg B$	
57		B	$\neg A$	A, $\neg B$	
58		$\neg B$	A	$\neg A$, B	
59		$\neg B$	$\neg A$	A, B	
60		A	$\neg A$, B	$\neg B$	
61		A	$\neg A$, $\neg B$	B	
62		$\neg A$	A, B	$\neg B$	
63		$\neg A$	A, $\neg B$	B	
64		B	$\neg A$, $\neg B$	A	
65		B	A, $\neg B$	$\neg A$	
66		$\neg B$	$\neg A$, B	A	
67	Type-3	$\neg B$	A, B	$\neg A$	
68	ranking	A, B, $\neg B$	$\neg A$		
69		$\neg A$, B, $\neg B$	A		
70		A, $\neg A$, B	$\neg B$		
71		A, $\neg A$, $\neg B$	B		
72		A	$\neg A$, B, $\neg B$		
73		$\neg A$	A, B, $\neg B$		
74		B	A, $\neg A$, $\neg B$		
75		$\neg B$	A, $\neg A$, B		

Table 1. Overview of possible rankings that a participant might give and their categories. Logically possible rankings are consistent with the Complement rule. Type-1, Type-2, and Type-3 rankings violate the Complement Rule. In a Type-1 ranking, two complementary events receive the first two places (or first place if there is a tie in the first place). In a Type-2 ranking, events that are not complementary to each other receive the first two places (or first place if there is a tie in the first place), but when an event receives the first place, its complement does not receive the last place. A Type-3 ranking can be categorized as both a Type-1 ranking and a Type-2 ranking.

Hypotheses The experimental manipulation resembles the setting in the simulation, since our objective is to test if the qualitative patterns we derived from the simulation hold empirically.

In the simulation, we manipulated inherent beliefs by taking different types of event sets as inputs. Specifically, we constructed three types of event sets:

- *edge event sets* consisting of four events with probabilities close to the endpoints of the probability scale,
- *middle event sets* consisting of four events with probabilities close to the midpoints of the probability scale, and
- *mixed event sets* consisting of two events with probabilities close to the endpoints of the probability scale and two events with probabilities close to the midpoints of the probability scale.

We also varied sample size and error rates. Figure 1 shows the simulation results. From these (i.e., visual inspection of Figure 1) we derive the following hypotheses:

- The probability of giving a logically possible ranking is highest for the mixed event sets, second-highest for the edge event sets, and lowest for the middle event sets, in the condition where ties are allowed.

- The conditional probability of giving a Type-3 ranking given that the participant does not give a logically possible ranking is highest for the middle event sets, second-highest for the mixed event sets, and lowest for the edge event sets, in the condition where ties are allowed.

- The conditional probability of giving a Type-1 ranking given that the participant does not give a logically possible or a Type-3 ranking is highest for the middle event sets, second-highest for the mixed event sets, and lowest for the edge event sets, in the condition where ties are allowed.

- The probability of giving a logically possible ranking is highest for the mixed event sets, second-highest for the edge event sets, and lowest for the middle event sets, in the condition where ties are not allowed.

- The condition probability of giving a Type-1 ranking given that the participant does not give a logically possible is highest for the middle event sets, second-highest for the mixed event sets, and lowest for the edge event sets, in the condition where ties are not allowed.

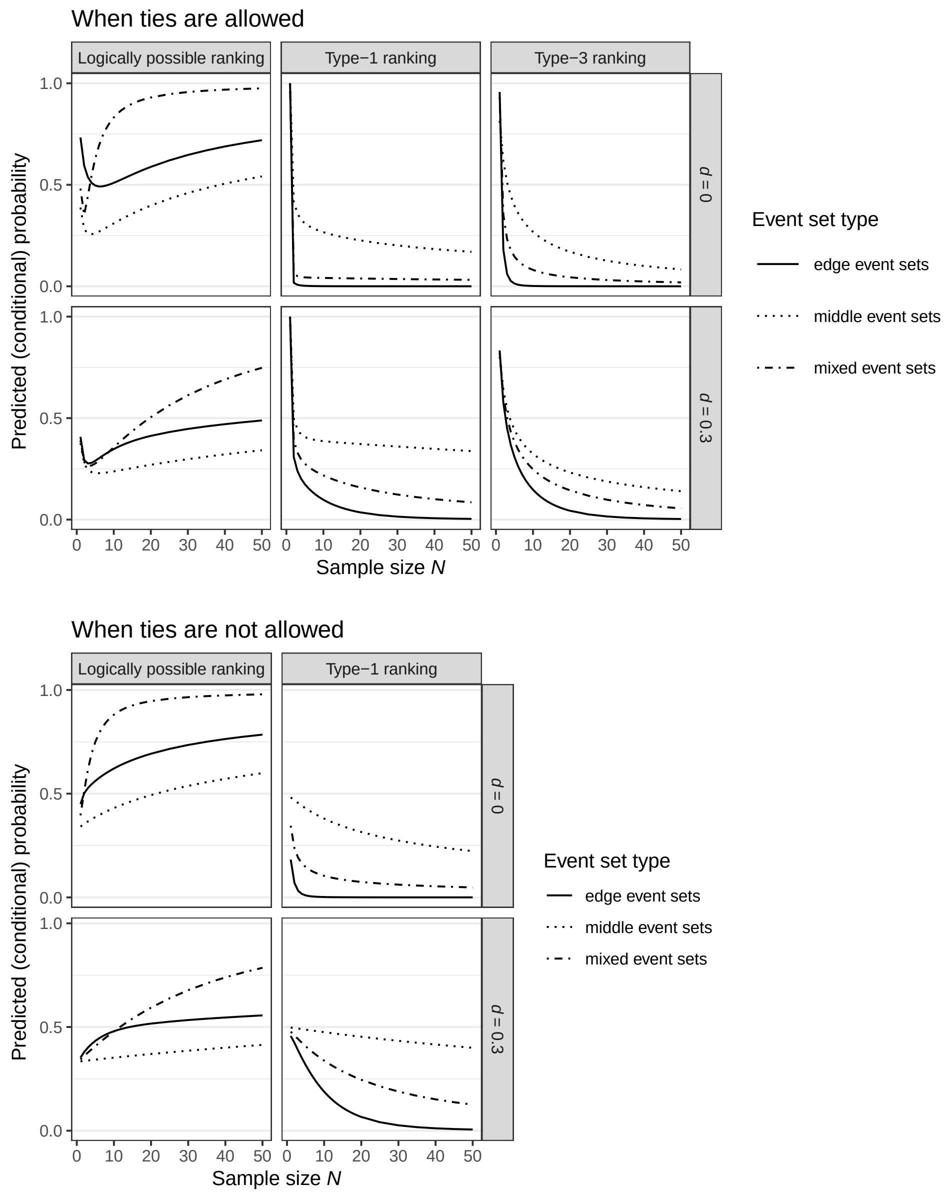

Figure 1. Predicted (conditional) probability of giving different types of rankings as a function of event set types, error rates d, and sample sizes N. Top row panels show the predictions for the condition where participants are allowed to give ties. Bottom row panels show the predictions for the condition where participants are not

allowed to give ties. The predicted (conditional) probabilities for each plot are as follows. "Logically possible ranking": probability of giving logically possible versus all other rankings. "Type-3 ranking": conditional probability of giving Type-3 versus Type 1 and Type 2 ranking. "Type-1 ranking": conditional probability of giving Type 1 versus Type 2 rankings.

Design Plan

Study type

Experiment. A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.

Blinding

No blinding is involved in this study.

Study design

We have a 2 (whether ties are allowed, between Ss) x 3 (event set type, within Ss) mixed-factorial design.

We manipulate between subjects whether participants are allowed to give ties when creating rankings for event sets. We manipulate within-subjects the type of event set that participants see. Each participant will be asked to provide rankings for three types of event sets: (1) edge event set, (2) middle event set, and (3) mixed event set.

In more detail, each participant sees 12 event sets: four edge event sets, four middle event sets, and four mixed event sets. Each event set consists of two pairs of complementary events. For example, for the event set $\{A, \neg A, B, \neg B\}$, A and $\neg A$ are a pair of complementary events, and B and $\neg B$ are another pair of complementary events. Positive events in two event pairs determine the event set type. Please note that in each pair of complementary events, there is one event containing negative words in its description, such as "not" or "no"; such events are referred to as negative events (noted as $\neg A$ or $\neg B$). The other event does not contain any negative words in its description; such events are referred to as positive events (noted as A or B).

To obtain the different types of events needed in this experiment, we performed a pilot study in which participants were asked to judge the subjective probability of events on a probability scale from 0% to 100%. From the pilot study, we selected

6 positive events with subjective probabilities close to 0, 6 positive events with subjective probabilities close to 1, and 12 positive events with probabilities close to 0.5. We refer to the event pair in which the positive event has a probability close to 0 as implausible event pairs. We refer to the event pair in which the positive event has a probability close to 1 as plausible event pairs. We refer to the event pair in which the positive event has a probability close to 0.5 as indifferent event pairs. All events in the pilot study were selected to have low variance in subjective probability.

We constructed the event sets using the events from the pilot study. As mentioned before, participants see four edge event sets. One of them is composed of two plausible event pairs, and one is composed of two implausible event pairs. The remaining two edge event sets are composed of one plausible event pair and one implausible event pair. For the four middle event sets that participants see, each of them is composed of two indifferent event pairs. As for the four mixed event sets, two of them are composed of one plausible event pair and one indifferent event pair. The other two are composed of one implausible event pair and one indifferent event pair.

Randomization Which of the events is shown in which event pairs is randomly determined anew for each participant. The order of the 12 event sets is also randomized anew for each participant.

Sampling Plan

We will use Prolific (https://www.prolific.co) to recruit 300 German participants. We will equally distribute the participants between two between-subject conditions. Thus, 150 participants will be assigned to the condition where ties are not allowed in the rankings, and 150 participants will be assigned to the condition where ties are allowed in the rankings.

Existing data **Registration prior to creation of data**. As of the date of submission of this research plan for preregistration, the data have not yet been collected, created, or realized.

Data collection procedures	Participants will be recruited using Prolific (https://www.prolific.co). Participants will be paid £2 for their participation. Participants must be at least 18 years old. Participants must currently live in Germany and have German as their first language.
Sample size	Our target sample size is 300 participants.
Sample size rationale	We have conducted a study previously which resembles the current study except for levels in within-subject manipulation. Specifically, in the previous study, we had a 2 (whether ties are allowed, between Ss) x 2 (event set type, within Ss) mixed-factorial design. We recruited 177 participants, 91 of whom were assigned to the condition where ties were allowed in the rankings and 86 of whom were assigned to the condition where ties were not allowed in the rankings. Each participant was asked to provide rankings for six edge event sets and six middle event sets. To get a comparable total number of observations, we thus need 150 ($\approx 91 \div 4 \times 6$) participants for each between-subject condition. Therefore, we need a total of 300 participants.
Stopping rule	We will stop data collection, once data from 300 participants is collected.

Variables

Manipulated variables	We manipulate between subjects whether participants are allowed to give ties when creating rankings for event sets. We manipulate within-subjects the type of event set that participants see. Each participant will be asked to provide rankings for three types of event sets: (1) edge event set, (2) middle event set, and (3) mixed event set.
Measured variables	When ties are prohibited in the rankings, there are in total 24 possible responses (i.e., rankings) over four events. We divide 24 rankings into three groups (see Table 1 in the Description section for details). The dependent variable will be the frequency of giving each category of responses (i.e., rankings).

When ties are allowed in the rankings, there are in total 75 possible responses (i.e., rankings) over four events. We divide 75 rankings into four groups (see Table 1 in the Description section for details). The dependent variable will be the frequency of giving each category of responses (i.e., rankings).

Analysis Plan

Statistical models

We have developed a multinomial processing tree (MPT) model for each between-subject condition. We will separate the data generated from two between-subject conditions and analyze them using the corresponding MPT model using a hierarchical-Bayesian approach.

MPT models account for categorical data following a multinomial distribution. However, unlike most MPT models in psychological research, MPT models we use here do not relate the probabilities underlying observed categorical responses to latent cognitive processes. Our goal is to use MPT modelling to decompose the multinomial distribution of rankings into a set of unconditional and conditional probabilities of responses.

In the following we present the model equations. First, we present the MPT model we develop for the condition where ties are allowed:

- Pr(logically possible ranking) $= l$,
- Pr(Type-3 ranking | not logically possible ranking) $= (1 - l) \times t_3$,
- Pr(Type-1 ranking | not logically possible ranking & not Type-3 ranking) $= (1 - l) \times (1 - t_3) \times t_1$,
- Pr(Type-2 ranking | not logically possible ranking & not Type-3 ranking) $= (1 - l) \times (1 - t_3) \times (1 - t_1)$.

Secondly, we present the MPT model we develop for the condition where ties are not allowed:

- Pr(logically possible ranking) $= l$,
- Pr(Type-1 ranking | not logically possible ranking) $= (1 - l) \times t_1$,
- Pr(Type-2 ranking | not logically possible ranking) $= (1 - l) \times (1 - t_1)$.

We will fit the MPT model jointly for each within-subject condition using a hierarchical-Bayesian approach (Klauer, 2010). Separate sets of population-level parameters will be estimated for each within-subject condition. The model fitting will be implemented via TreeBUGS Package (Heck, Arnold, & Arnold, 2018).

To test the hypotheses, we will assess if there are statistically meaningful differences among group-level parameter estimates obtained in different within-subject conditions. In the condition where ties are allowed, we will assess the differences in three group-level parameters: parameter l, t_1 and t_3. If ties are not allowed, we only need to examine two group-level parameters: parameter l and t_1.

For parameter l, we first calculate a difference distribution by subtracting the MCMC posterior samples we get in the edge event sets condition from the MCMC posterior samples we get in the mixed event sets condition. We then calculate a difference distribution by subtracting the MCMC posterior samples we get in the middle event sets condition from the MCMC posterior samples we get in the edge event sets condition. Lastly, we calculate a difference distribution by subtracting the MCMC posterior samples we get in the middle event sets condition from the MCMC posterior samples we get in the mixed event sets condition.

For parameters t_1 and t_3, we need a different set of difference distributions. Firstly, we calculate a difference distribution by subtracting the MCMC posterior samples we get in the mixed event sets condition from the MCMC posterior samples we get in the middle event sets condition. We then calculate a difference distribution by subtracting the MCMC posterior samples we get in the edge event sets condition from the MCMC posterior samples we get in the mixed event sets condition. Lastly, we calculate a difference distribution by subtracting the MCMC posterior samples we get in the edge event sets condition from the MCMC posterior samples we get in the middle event sets condition.

Transformations Enter your response here.

Inference criteria We will conclude that there is a statistically meaningful difference between the group-level parameter estimates obtained from two with-subject conditions if more than 95% of the probability mass of the difference distribution is above 0.

Data exclusion

We will include an attention check item to determine eligibility for inclusion. Participants that do not pass the attention check item will not be included in the analysis.

In the attention check item, we will ask participants to provide a ranking for an event set consisting of the below four events:

- An einem zufällig ausgewählten Tag im Jahr 2020 war Berlin die Hauptstadt Deutschlands.
- An einem zufällig ausgewählten Tag im Jahr 2020 war Berlin NICHT die Hauptstadt Deutschlands.
- In einem zufällig ausgewählten Krankenhaus in Deutschland wird das nächste Neugeborene ein Mädchen sein.
- In einem zufällig ausgewählten Krankenhaus in Deutschland wird das nächste Neugeborene KEIN Mädchen sein.

Missing data

Participants that do not complete the study will be excluded.

Exploratory analyses (optional)

Enter your response here.

Other

Other (Optional)

Enter your response here.

References

Costello, F., & Watts, P. (2014). Surprisingly rational: Probability theory plus noise explains biases in judgment. Psychological Review, 121(3), 463–480. https://doi.org/10.1037/a0037010

Heck, D. W., Arnold, N. R., & Arnold, D. (2018). TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling. Behavior Research Methods, 50(1), 264–284. https://doi.org/10.3758/s13428- 017-0869-7

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. Psychometrika, 75(1), 70–98. https://doi.org/10 .1007/s11336-009-9141-0